

Mitä Vinogradov ei voinut arvata

Instrumentarii rusistiki: Korpusnyje podhody. Red. Arto Mustajoki, Mihail Kopotev, Leonid Birjulin & Jekaterina Protassova. Helsinki: Department of Slavonic and Baltic Languages and Literatures, 2008. 371 s.

Laajoihin sähköisiin tekstiaineistoihin perustuvasta korpuslingvistiikasta on tullut keskeinen tutkimusala ja yksi kielen- ja kääntämisen tutkimuksen suosituimmista lähestymistavoista. Venäjän kielen tutkimuksen instrumentteja esittelevien julkaisujen teeman Slavica Helsingiensia-sarjassa aloittaa kieliteknologian kehityksestä ammentava *Rusistikan välineistö: Korpusmenetelmät*. Teokseen on kirjoittanut 23 venäjän kielen tutkijaa Suomesta, Venäjältä ja kauempaakin. Kahta englanninkielistä artikkelia lukuun ottamatta teos on venäjänkielinen.

Johdantoartikkelissa käsitellään yleisesti korpuslingvistiikkaa ja korpus tutkimuksen nopeaa kehitystä. Suurin osa kirjoittajista on käyttänyt aineistonaan hieman lähdössä myöhästynyttä mutta siten nopeasti edennyt, kielentutkimuksen tarpeisiin kehitettyä *Venäjän kielen kansalliskorpusta* tai *Integrum*-tietokantajärjestelmää, johon päivittyvä jatkuvasti lähes kaikki medioiden tuottama teksti (yli 400 miljoonaa julkaisua, kun useat korpuukset eivät pääse edes samaan määrään sanoja).

Korpusmenetelmien mahdollisuuksia testaavat M. Hohlova, joka vertailee kollokaatioiden eli vakiintuneiden sanayhdistelmien tilastollisia tunnistamismenetelmiä, ja Galina Nikiporets-Takigawa etsiessään uudissanoja metalingvististen lausekkeiden

avulla (”kuten nykyään on tapana sanoa” yms.). Viimemainittu on myös toimittanut teoksen *Integrum: eksaktit menetelmät ja humanistiset tieteet* (2006). Mihail Kopotev on tutkinut venäjän sijajärjestelmää esimerkkinä siitä, miten voitaisiin luoda (taajuussanakirjan tapaan) kielipiillisten kategorioiden yleisyyden huomioon ottava kvantitatiivinen kielioppi. Mihail Mihailovin katsaus ’älyyn’ liittyvästä sanastosta 1800-luvulta nykypäivään pyrkii osoittamaan lemmatun ja merkityn tekstikorpuksen arvon kielen kehityksen seuraamisessa.

Kriittistä diskurssianalyysia edustaa L. Ryazanova-Clarke, joka paljastaa, miten Putin, toimittajat ja kansa rakentavat kansallista identiteettiä suorassa radio-ohjelmassa. Korpus aineistolähteenään kirjoittaja analysoi syvälliseksi ja spontaaniksi mainostettua lähetystä, jossa kysyjistä muutama promillen sadasosa pääsi edustamaan kansaa. Kirjoitus päättyy sitaattiin spektaakkeliyhteiskunnasta, joka ”on integroitunut todellisuuteen samassa määrin kuin se kuvaa sitä ja [...] rekonstruoit sitä sellaiseksi kuin se kuvaa sitä” (s. 330). Jekaterina Protassovan artikkeli lähtee prototyyppisistä mielikuvista: venäläinen on laiska ja työskentelee huonosti, työnteossa on tärkeätä yhteisö ja toisaalta tehdään, mitä johtaja sanoo, säännöistä välittämättä. Kvantitatiivista argumentointia ei ole tehty mutta ”työskentelyverbien” kontekstin valossa ”venäjäksi työskentely tarkoittaa tuloksen ennakoimattomuutta, työskentelyä piittaamatta säännöistä, määrärajoista tai itsestään” (s. 300). Verbiin *rabotat* liittyy konnotaatio pakonomaaisesta tai

vastentahtoisesta työstä, toisin kuin verbiin *truditsja*. I.T. Vepreva tutkii, mitkä sanat tai ilmaisut ja miksi herättävät kansalaisissa ärtymystä. Synnän voi olla – paitsi luotaantyöntävä tarkoite sinänsä (’sota’, ’lapsirikollisuus’) – myös tyytymättömyys epätarkkaan tai tyyliin sopimattomaan sanankäyttöön: ilmaisun ”esteettinen” vastenmielisyys, kielteinen suhtautuminen muoti-ilmaisuihin tai ideologisesti värityneeseen sanastoon. Toisaalta kaivataan kiertoilmaisuja, kansa vaatii poliittista korrektiutta. Marjatta Vanhala-Aniszewski ja Lea Siilin ovat analysoineet sanaa *veliki* ’suuri’ luonnehtimassa maata, kieltä ja kulttuuria. Sana kuuluu uusimpaankin venäläiseen diskurssiin eivätkä suuren ja mahtavan tarkoitteetkaan ole juuri muuttuneet. Sen sijaan 90-luvun pessimismi tai haikailu menneen suuruuden perään on nykyteksteissä väistymässä ja uusi uho heräämässä.

Suosituksi tutkimuskohteeksi ovat korpus tutkimuksen myötä tulleet useamman sanan (fraasiutuneet) sanayhdistelmät, joita sanakirjat eivät ”nominatiivihakuisina” yleensä esitä. T.I. Kustova käsittelee ’joka tapauksessa’ -tyyppisiä rakenteita (*vo vsjakom slutsaje*, vrt. *v každom slutsaje* ’jokaisessa tapauksessa’). Fraaseologismeiksi hän katsoo myös sanaliitot tyyppiä *železnaja doroga* ’rautatie’, vaikka voisi puhua leksikaalistuneista idiomeista. Selvimmin konstruktiokieliopin ajatuksiin tukeutuvaa tutkimusta edustaa A.-J. Kyröläisen artikkeli, jossa lähdetään rakenteen muodon ja merkityksen erottamattomasta kokonaisuudesta eikä kuten aiemmissa tutkimuksissa verbin merkityksestä. Teoreettisena lähtökohtana on pelkän frekvenssin kyseenalaistaminen rakenteen kieliopillistumisen edistäjänä,

esimerkkinä persoonattomat datiivirakenteet ja suljetun luokan muodostavat liikeverbit (*komul. /ne/ bežitsja* '(jonkun) tekee /ei tee/ mieli juosta'). Rakenteen produktiivisuudesta saadaan näyttöä Yandexista ja Googlesta, joista löytyy esiintymiä myös Integrumissa negatiivisen hakutuloksen antaneista verbeistä. O.N. Ljaševskaja ja E.V. Rahilina kuuluttavat tutkimusta syntaktisten rakenteiden lainautumisesta sanan mukana. Heidän mukaansa diskurssissa ei ole monimerkityksisyyttä eikä kielenpuhuja huomaa käyttävänsä samaa sanaa eri tavoin. Syntaksiltaan identtissä rakenteissa *u Ya problemy s X-om* 'Y:llä on ongelmia X:n kanssa' ongelman aiheuttajina ovat kaunokirjallisisissa teksteissä esineet (auto, jarrut jne.), ruumiinosat ja elimet (sydän, iho) sekä henkilönnimet (johtaja, kävijä); lehtiteksteissä teonnimet (siivous, maksu). Onko kyse puutteesta, sairaudesta, vammasta tai viasta, määräytyy X:n semanttisesta luokasta riippumatta rakenteen subjektista (kokijasta). E.R. Dobrušina artikkelin otsikossa "Ylistysvirrestä dogmaattiselle altruismille" on tavoiteltu lauseketta sanoista, joita artikkelissa käsitellään (*akafist* 'ylistysvirsi, -laulu, akatisti', *altruist*, *dogma*). Vaikka uskonnonharjoitukseen liittyvät (kreikkalaisperäiset) oppisanat eivät suomessa juurikaan ole muussa käytössä, niin suomenkin "ylistysvirsi" muuhun kuin Jumalan ylistämiseen liittyessään sisältäneen negatiivisen arvotuksen.

N.R. Dobrušina ja S.A. Pantelejeva käsittelevät venäjän ns. kollektiivilukusanonoja (vrt. 'kolmikko' < *kolme*), joilla heidän havaintojensa mukaan on yksilöivä funktio. Aineiston valossa kollektiiviset lukusanat esiintyvät käytännössä vain nominatiivissa ja genetiiivissä ja niitä käytetään yhdessä toimiviksi mielletä-

vistä yksilöistä, esim. sotilasta tai työntekijöistä (mutta ei johtajista tai upseereista). T.V. Šmeljovan artikkeli pohtii, miksi Vinogradovin ennuste (1947) omistusadjektiivien katoamisesta ei näytä toteutuvan. Nämä adjektiivit näyttävät säilyvän ja lisääntyvän erilaisissa, myös yritysten ja tuotteiden nimissä. Käyttö perhepiirissä sukulais- ja henkilönnimistä (mamin 'äidin', *Tanin* 'Tanjan') tuskin on uusiokäyttöä, vaan katkeamatonta perinnettä. Substantiivisen ja nykyisen norminmukaisen adjektiivitaivutuksen sekaantumista on havaittavissa. Mutta Vinogradov ei mitenkään voinut arvata, että tulee aika, jolloin paljon siitä, mikä "huolettisesti karkotettiin", palaa takaisin. E.A. Grišina paneutuu *vo*-partikkelin varianttien (*vo*, *vo*, *o*, *ot*) merkityseroihin vapaassa puheessa ja ehdottaa sille myös radikaalisti perinteisestä poikkeavaa etymologiaa: huudahduksen ja partikkelin homonymia olisi johtanut proteettisen *v*:n lisäykseen jälkimmäisessä.

Olga Pussinen on Arto Mustajoen kanssa tutkinut *po*-etuliitteellä muodostettujen verbien ekspansiota. Sanaryhmän diskurssimerkityksen tulkinta ei ehkä ole artikkelin vakuuttavin osa, mutta sen etsintä on motivoitua. Suurinta osaa näistä verbeistä ei löydy sanakirjoista. Artikkelin merkittävä myös ensimmäisenä laajana tekstikorpusta hyödynnettävänä tämän kysymyksen tutkimuksena niukassa korpuspohjaisessa aspektitutkimuksessa. I. Kor Chahine tutkii huudahduksen kaltaisia, useimmiten yksitavuisia ja ääntä jäljitteleviä sanoja, joita vastaa usein samankantainen verbi. Kor Chahinen näille *molskis*-tyyppisille predikaatin korvikkeille ehdottama termi "narratiivinen predikatiivi" perustuu havaintoon, että kyseiset sanat vievät kertomusta eteenpäin. Useimmissa muissa ehdo-

tuksissa (esim. *Garden prédicat interjective*) on mukana viittaus huudahdukseen.

I.V. Dubrovina tarkastelee murhetta ja ikävää ilmaisevien verbien semantiikkaa: tuntemuksen syyhyin viittaava epäsuora objekti esiintyy joillakin verbeillä vain *o-* (*tužit* 'murehtia'), joillakin sekä *po-* että *o-*preposition kanssa (*toskovat* 'kaivata'). Variaation syyksi Dubrovina esittää, että 'sureminen' (*o-*preposition kanssa) on tahdonalaisempaa kuin 'kaipuu' *po-*prepositiota käytettäessä. Ljaševskaja tarkastelee ilmaisia tyyppiä *videt* (*svoimi*) *glazami* 'nähdä (omin) silmin', *dumat* (*golovoj*) 'ajattella (päällään)' sen valossa, että tyyppillisesti ihminen näkee silmillään ja ajattelu tapahtuu päässä, ja sen vuoksi instrumentin ilmi paneminen on tautologista. "Tautologinen ekskorporaatio" on kuitenkin systemaattinen ilmiö korpusmateriaalissa ja sillä on semanttinen motivaatio. T.I. Steksova tutkii verbin *delat* 'tehdä' muodostamia, kahta subjektia edellyttäviä kausatiivirakenteita. Venäjän rakenteiden *delat kogo tšem* 'tehdä joku joksikin' ja *delat iz-kogo tšto* 'tehdä jostakusta jokin' ero on samantapainen kuin suomen lähes synonyymisten "tulosrakenteiden" (joku tulee raskaaksi tai jostakin tulee raskas). Tapauksessa, jossa 'ministereistä haluttiin tehdä rikollisia', *iz-*rakenne voidaan tulkita siten, että 'ministerit haluttiin esittää rikollisina' (artikkelissa tulkinnot on esitetty väärin päin).

Monipuolisen ja tervetulleen kokoelman teknisessä yhtenäisyydessä voi nähdä pieniä puutteita, kun tekstien muodollinen viimeistely on jätetty yksittäisten kirjoittajien vastuulle. Painovirheet ja lipsahdukset olisi pitänyt siistiä pois, lisäksi kirjoittajien esittelyä, edes sähköpostiosoitteita, jää kaipaamaan. Kun kirjoittajien etunimiä ei mainita, sukupuolikin voi

jäädä epäselväksi. Vaikka saahan se kenties jäädäkin, kun kirjoitan suomeksi... mutta jospa joku

kirjoittaa muulla kielellä?

Hannu Tommola

Metodologiaa ja ideologiaa

Hannu Kemppanen: *Avainsanoja ja ideologiaa: Käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Joensuu: Joensuun yliopisto, 2008.

Hannu Kemppasen väitöskirjan otsikko on virkistävän konstailematon ja kertoo mistä työssä on kyse. Sama selkeys ja selittämisen perusteellisuus leimaa koko kirjaa. Johdantoluvussa tekijä esittelee tutkimuksen taustaa, tavoitteita, aineistoja ja metodeja. Hän erittelee korpuslingvistisen käännöstutkimuksen perusoletuksia ja lähtökohtia. Tutkimuksen kannalta keskeinen termi on ”ideologia”. Tekijä toteaa käsitteen olevan huonosti määritelty ja mainitsee omaksi lähtökohdakseen ideologian läsnäolon kaikkialla. Hän myös esittelee ideologisen funktion käsitteen. Kemppanen kartoittaa johdantoluvussa myös käännöstutkimuksen historiaa siltä osin mikä liittyy hänen edustamaansa tutkimussuuntaukseen. Tämä asiantuntevan kriittinen kuvaus johdattelee lukijan luontevasti ko. tutkimuksen lähtökohtiin.

Toisessa luvussa Kemppanen kertoo ideologian käsitteestä ja sen käyttämisestä käännöstutkimuksessa. Perusteet ideologia-käsitteen käytölle tulevat hyvin esiin. Tekijä kirjoittaa aineistona olevien historiatekstien luonteesta, mainiten historiatekstien genren olevan erikoislaatuinen, koska se on vain osittain tiettyä tarkoitusta varten olevaa asiategstiä, jossa on osin kaunokirjallisuudelle tyypillisiä piirteitä. Tekijän mukaan historiatekstit ovat erityisen kiinnostavia myös

siksi että niissä ideologiset piirteet ovat selvästi esillä. Tämä johtuu siitä että historiateksti on tiukasti sidoksissa aikaan ja paikkaan ja niihin liittyviin arvoihin ja politiikkaan.

Kemppanen ilmoittaa tutkimuksen ensisijaisen tavoitteen olevan metodologinen. Analyysimenetelmistään kertoessaan hän esittelee ”työpakki-ajattelua”, jonka mukaan saman ilmiön eri puolia voi valottaa erilaisia työkaluja käyttäen. Analyysi muodostuu näin vaiheittaiseksi. Työpakin työkalut ovat analyysissa käytettäviä ohjelmia ja niihin liittyviä teoreettisia malleja. Käytännönläheinen metafora korostanee sitä, että yhdentyyppisellä analyysi- ja selitysmallilla kieliaineistosta pääsee yhdenlaisiin tuloksiin, joita toisentyypinen analyysi voi hyvin täydentää.

Kolmannessa luvussa Kemppanen käsittelee avainsana-analyysin metodia ja sen toteuttamista sekä teoreettisesta että käytännön näkökulmasta. Hän esittelee avainsana-termin historiaa ja sen avulla tehtyjen tutkimusten tuloksia. Kemppanen käsittelee monipuolisesti niitä etuja, joita korpuslingvistiikan mahdollisuudet tuovat avainsana-analyysiin. Hänen mukaansa tutkimus ”pyrkii kehittämään avainsana-analyysia nimenomaan menetelmänä, jolla voidaan analysoida eri tekstijoukkojen välisiä suhteita” (s. 57). Tutkimuksen metodologia on jälleen selitetty yksityiskohdittaisesti, myös tekniseltä kannalta. Vertailtavina ovat käännetyt ja ei-käännetyt historiatekstit, joiden valintakriteerit on selitetty seikka-peräisesti. Kemppanen käsittelee myös lemmatun ja lemmaamat-

toman aineiston hyviä ja huonoja puolia antaen esimerkkejä. Esitellässään käyttämänsä WordSmith Tools -konkordanssiohjelmaa hän on jopa ottanut mukaan runsaasti esimerkkejä siitä miten ohjelma konkreettisesti näyttää tuloksia. Nämä kuvat ovat arvokas lisä tekstin antamaan informaatioon. Kemppainen erottaa toisistaan kollokaation, klusterin ja idiomien, ja pohtii laajemminkin sanojen yhdessäesiintymistä ja sen teoreettisia tulkintoja.

Neljännessä luvussa Kemppanen käsittelee historiatekstien avainsana-analyysin tuloksia. Hän raportoi tutkimuksen tuloksia analyysivaiheiden mukaisesti ja antaa kunkin analyysivaiheen tulokset erikseen. Kemppasen mukaan menettelyllä on ”metodologista tai ehkä jopa pedagogista merkitystä, koska olettamusten kirjaaminen esiin antaa kuvan siitä, kuinka korpusmateriaalista nousevia tietoja voidaan mahdollisesti lukea ja tulkita” (s. 92). Esimerkiksi käsitellessään sanan *Suomi* kollokaatteja Kemppanen osoittaa vaihe vaiheelta taulukoiden avulla miten analyysi on tehty, ml. mitä analyysiin vaikuttaneita välitulkintoja tutkija on tehnyt ja miten nämä tulkinnot ovat vaikuttaneet tutkimuksen kulkuun. Samalla tavalla hän esittelee eri sanojen frekvenssejä teksteissä ja esittää tulkintoja siitä millä näitä eroja voi selittää. Usein toistuva selitys on lähdekieleessä olevaan stimulukseen reagoiminen, mikä selittää samojen sanojen ja sanontatapojen erilaista frekvenssiä käännöksissä ja ei-käännöksissä. Kemppanen myös vertailee tuloksiaan siihen, mitä aiempien tutkimusten nojalla tiedetään samojen sanojen frekvensseistä.

Kemppasen kantava idea hyödyntää erilaisia metodeja työpakki-ajattelunsa mukaisesti tulee esille myös siinä, miten hän ei vain selitä tarkkaan tutkimuksensa vaiheita ja metodeita, vaan myös antaa vaihtoehtoja siitä,